

IMPLEMENTASI ANALISIS CLUSTER BERTIERARKI PADA BUNGA MAWAR

Muhamad Azhari

Program Studi Ilmu Komputer FMIPA Univ. Mulawarman

ktob34@yahoo.com

Abstrak - Analisis komponen utama merupakan salah satu metode yang digunakan untuk mereduksi variabel peubah ganda yang besar jumlahnya. Informasi direpresentasikan menjadi variabel-variabel baru pada metode tersebut. Analisis Komponen Utama yang dipadukan dengan analisis cluster dapat digunakan untuk mengelompokkan sekumpulan obyek berdasarkan ukuran kemiripannya. Pada implementasi analisis cluster bertierarki pada bunga mawar dengan metode perhitungan jarak euclidean menghasilkan hasil eigenvalue λ_3 , λ_5 dan λ_8 . Masing-masing eigenvalue tersebut mampu memberikan informasi sebesar 78%, 87,9% dan 93,8%. Berdasarkan hasil tersebut dilakukan analisis cluster bertierarki dengan Metode Single Linkage, Complete Linkage, Average Linkage dan Wards Linkage

Kata Kunci : Analisis Komponen Utama, Cluster bertierarki, Jarak Euclidean

Metode analisis komponen utama (PCA) merupakan salah satu teknik yang digunakan untuk menentukan komponen utama dari suatu distribusi. Metode tersebut diterapkan pada vektor-vektor karakteristik (*eigenvectors*) dari matriks kovariansi dari kumpulan citra bunga mawar yang telah dinormalisasi. Vektor-vektor karakteristik tersebut merupakan kumpulan *feature* yang menjadi karakteristik bersama dari citra bunga mawar.

Pada pengenalan pola masing-masing variasi pada citra bunga mawar dapat direpresentasikan dengan tepat dalam suatu kombinasi linier karakteristik (*eigenfaces*). Variasi paling berarti dapat direpresentasikan melalui vektor dengan nilai karakteristik (*eigenvalues*) terbesar.

Analisis *cluster* adalah salah satu teknik pengelompokan n obyek ke dalam m kelompok ($m \leq n$). Pengelompokan dapat dilakukan berdasarkan pada ukuran kemiripan (*similarities*) atau ketidakmiripan (*dissimilarities*). Pengelompokan bertujuan untuk menemukan pengelompokan alamiah (menghindari keragaman obyek pengamatan dalam satu kelompok lebih homogen daripada keragaman antar kelompok). Hasil pengelompokan berguna dalam pendugaan,

pengidentifikasian pencilaan dan penentuan hipotesis awal [1].

Ada dua metode pengelompokan yaitu metode bertierarki dan tidak bertierarki. Metode bertierarki digunakan bila jumlah kelompok yang diinginkan tidak diketahui. Sedangkan metode tidak bertierarki digunakan bila jumlah kelompok yang diinginkan diketahui. Proses hirarki adalah konsep jarak antara suatu obyek dengan sebuah kelompok dan jarak antara dua kelompok.

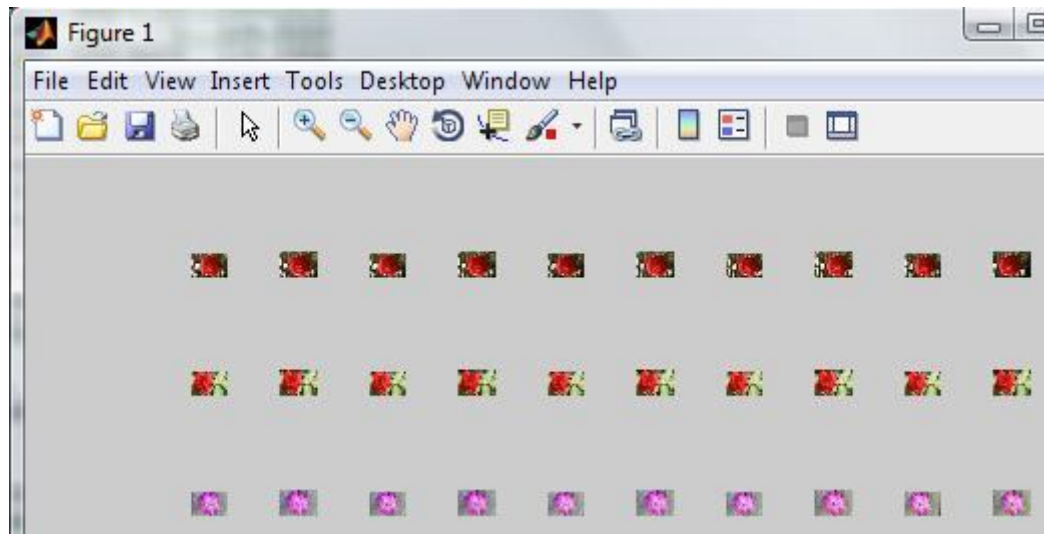
Ukuran jarak yang biasa digunakan adalah jarak *Euclid* dan *Mahalanobis*. Jarak *Euclid* digunakan jika informasi mengenai sebaran data asal tidak diketahui dan peubah-peubah yang diamati tidak berkorelasi [2].

METODOLOGI

1. Data

Penelitian menggunakan data citra bunga mawar sebanyak 30 buah (40 x 60 piksel) yang berasal dari 3 jenis bunga mawar. Masing-masing citra tersebut diambil dengan pemotretan digital sebanyak 10 kali.

Berikut adalah visualisasi hasil pemotretan 30 gambar .



Gambar 1 Visualisasi citra bunga mawar

2. Pengolahan data

- Software MATLAB R2009a digunakan untuk mengkonversi masing-masing citra (40x60 piksel) ke bentuk vektor 2400x1 untuk selanjutnya digabungkan dengan citra lainnya sehingga menjadi matrik berukuran 2400x30.
- Software MINITAB untuk pengolahan data statistik antara lain untuk Analisis Komponen Utama dan Analisis *Cluster*.

3. Metode Analisis yang digunakan adalah

- Analisis Komponen Utama untuk membentuk variabel-variabel baru yang merupakan kombinasi linear dari variabel yang sudah ada.
- Analisis *Cluster*, yang bertujuan untuk mengelompokkan obyek berdasarkan ciri yang dimilikinya. Metode *Cluster* yang digunakan adalah Metode *Cluster* berhierarki (*Single Linkage*, *Complete Linkage*, *Average Linkage* dan *Wards*) dengan Metode Perhitungan Jarak *Euclidean*.

ANALISIS DAN PEMBAHASAN

1. Analisis Komponen Utama

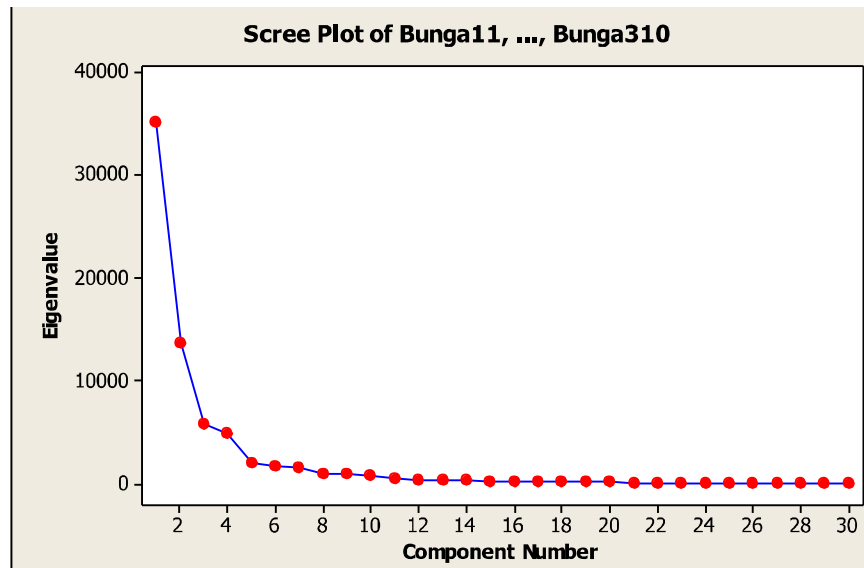
Pengolahan data yang dilakukan dengan Minitab mendapatkan hasil untuk analisis komponen utama sebagai berikut :

Tabel 1 Hasil Pengolahan Analisis Komponen Utama

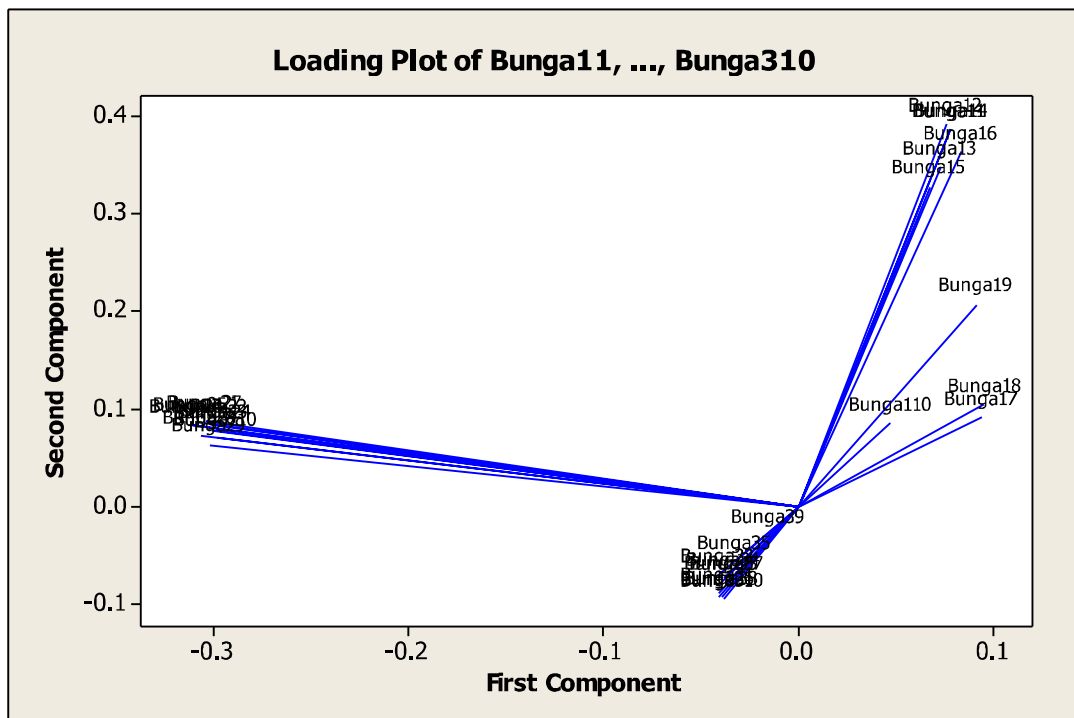
Variabel	Eigenvalue		
	Varians	% Proporsi	% Kumulatif
1	35210	0.503	0.503
2	13637	0.195	0.698
3	5754	0.082	0.78
4	4925	0.07	0.85
5	1980	0.028	0.879
6	1684	0.024	0.903
7	1472	0.021	0.924
8	949	0.013	0.938

Tabel 1 menunjukkan jika terdapat 3 variabel baru yang mempunyai variasi sebesar 5754, maka perbandingan nilai λ_3 dengan varians total adalah 0.082 (8,2%). Sehingga total informasi yang didapat dari λ_3 adalah 78 %. Informasi dari variabel baru yang mampu mewakili keseluruhan (90% s.d 95%) dihasilkan dengan menggunakan λ_5 dan λ_8 .

Sebanyak 87% informasi dengan varians sebesar 1980 dapat diterangkan dengan menggunakan λ_5 . Sementara 93,8 % informasi dengan varians sebesar 949 dapat diterangkan menggunakan λ_8 .



Gambar 2 Scree plot Bunga 11 s.d. 310



Gambar 3 Loading Plot citra bunga mawar

Scree Plot pada Gambar 2 menunjukkan nilai *eigenvalue* dari yang terkecil sampai terbesar. Pada sumbu x menunjukkan jumlah angka yang sesuai untuk banyak komponen. Pengambilan nilai *eigenvalue* didasarkan pada belokan tajam yang menunjukkan perubahan yang signifikan dibanding titik-titik yang relative stabil. Dari Scree Plot di atas Nilai *Eigenvalue* ke-3 dapat disimpulkan sebagai titik belok yang tajam. Sedangkan nilai *eigenvalue*

ke-5 dan ke-8 diambil berdasarkan tingkat kepercayaan mendekati 85 % dan 95 %.

Pada Gambar 3 menunjukkan posisi jarak antar bunga, Himpunan bunga 1-10 lebih dekat jaraknya dengan bunga 21-30 daripada dengan Bunga11-20. Padahal dari segi warna bunga 1-10 sama-sama merupakan bunga mawar warna merah, hanya saja pada Bunga11-20 terlihat adanya daun yang cukup dominan sehingga mempengaruhi perhitungan.

Salah satu kombinasi linear yang dihasilkan oleh variabel baru dihasilkan PC3. Hasilnya adalah sebagai berikut:

PC3=0.172 Bunga11 + 0.162 Bunga12 + 0.153 Bunga13 + 1.137 Bunga14 + 0.147 Bunga15 + 0.088 Bunga16 - 0.504 Bunga17 - 0.535 Bunga18 - 0.271 Bunga19 - 0.139 Bunga20 - 0.048 Bunga21 - 0.046 Bunga22 - 0.045 Bunga23 - 0.049Bunga24 - 0.044 Bunga25 - 0.044 Bunga26 - 0.044 Bunga 27 - 0.025 Bunga28 - 0.023 Bunga 29 - 0.026 Bunga210 + 0.161 Bunga 31 + 0.159 Bunga32 + 0.162 Bunga33 + 0.158 Bunga34 + 0.133 Bunga35 +

0.141 Bunga36 + 0.161 Bunga 37 + 0.158 Bunga38 + 0.087 Bunga39 + 0.159 Bunga310

Dari Persamaan di atas jelas terlihat koefisien masing masing variabel terhadap variabel baru (PC3). Analisis selanjutnya (*cluster*), digunakan cluster sebanyak 3, 5 dan 8.

2. Analisis Cluster

Hierarki

Analisis *Cluster* dengan metode berhirarki (*Single, Complete, Average* dan *Ward*), hasil pengolahan secara lengkap dapat dilihat pada Lampiran

Tabel 2 Hasil Pengolahan Cluster dengan Metode berhirarki

Metode	Cluster =3	Cluster=5	Cluster = 8
Single	1. Bunga 1-10 2. Bunga 11 - 20 3. Bunga 21- 30	1. Bunga 1-6 2. Bunga 7-9 3. Bunga 10 4. Bunga 11 - 20 5. Bunga 21 - 30	1. Bunga 1 -6 2. Bunga 7 - 8 3. Bunga 9 4. Bunga 10 5. Bunga 11 -20 6. Bunga 21, 26, 28, 30 7. Bunga 22, 23, 24, 25, 27 8. Bunga 29
Complete	1. Bunga 1-10 2. Bunga 11 - 20 3. Bunga 21- 30	1. Bunga 1-6 2. Bunga 7-9 3. Bunga 10 4. Bunga 11 - 20 5. Bunga 21 - 30	1. Bunga 1 -6 2. Bunga 7 - 8 3. Bunga 9 4. Bunga 10 5. Bunga 11 -20 6. Bunga 21, 26, 28, 30 7. Bunga 22, 23, 24, 25, 27 8. Bunga 29
Linkage	1. Bunga 1-10 2. Bunga 11 - 20 3. Bunga 21- 30	1. Bunga 1-6 2. Bunga 7-9 3. Bunga 10 4. Bunga 11 - 20 5. Bunga 21 - 30	1. Bunga 1 -6 2. Bunga 7 - 8 3. Bunga 9 4. Bunga 10 5. Bunga 11 -20 6. Bunga 21, 26, 28, 30 7. Bunga 22, 23, 24, 25, 27 8. Bunga 29
Ward	1. Bunga 1-10 2. Bunga 11 - 20 3. Bunga 21- 30	1. Bunga 1-6 2. Bunga 7-10 3. Bunga 11 - 20 4. Bunga 21, 26, 28, 30 5. Bunga 22, 23, 24, 25, 27, 29	1. Bunga 1 -6 2. Bunga 7 - 8 3. Bunga 9 4. Bunga 10 5. Bunga 11 -20 6. Bunga 21, 26, 28, 30 7. Bunga 22, 23, 24, 25, 27 8. Bunga 29

Table 2 menunjukkan hasil peng-*cluster*-an dengan menggunakan empat metode yang berbeda hampir tidak ditemukan perbedaan. Namun pada metode *ward* yakni pada *cluster*=5 ditemukan hasil yang berbeda. Hal tersebut dikarenakan prosedur pengklasteran Wards yang mencoba meminimumkan informasi yang “*loss*” dari penggabungan 2 kelompok. Pengambilan jarak antara dua cluster didasarkan pada total *sum of square* dari dua cluster pada masing-masing variabel. Sehingga Ward dikatakan memakai pendekatan analisis varians untuk menghitung jarak antar kluster. Jumlah kuadrat dari setiap dua cluster yang dapat dibentuk berusaha dikecilkan. Disini, kelebihan Metode Ward dibanding metode lainnya adalah lebih efisien dan cenderung menciptakan kluster berukuran yang lebih kecil. Keseluruhan pengolahan *cluster* memperlihatkan bahwa Bunga 11-20 selalu terkumpul menjadi 1 *cluster*.

SIMPULAN

Penelitian menghasilkan simpulan:

1. Metode *cluster* sangat sensitive terhadap *outlier* atau *noise points*
2. Metode *cluster* berhierarki tidak dipengaruhi oleh sumber variasi dan kesalahan.
3. Tidak ada ketetapan dalam pemindahan obyek-objek yang kemungkinan pengelompokkannya tidak tepat pada metode *cluster* berhierarki.
4. Analisis *cluster* mengharuskan untuk melihat kecocokan objek-objek hasil peng-*cluster*-an.

DAFTAR PUSTAKA

- [1] Johnson R, Wicherin D. *Applied Multivariate Statistical Analysis*, Sixth Edition, Prentice Hall, New Jersey. 2007
- [2] Dillon WR, Goldstein M. *Multivariate Analysis Methods and Application*, John Wiley & Sons. New York. 1984
- [3] MINITAB Reference Manual
- [4] Using MATLAB. The MathWorks, Inc. 1984-1999